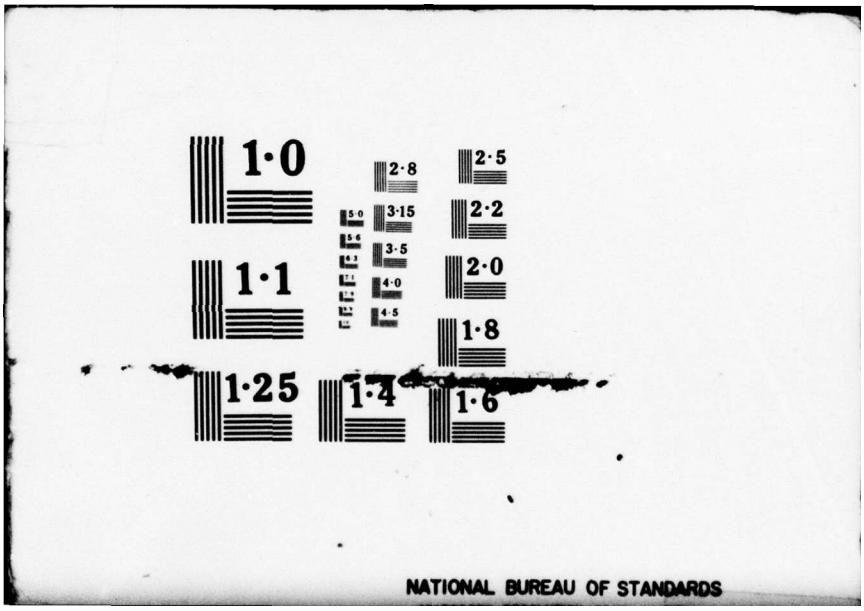
AD-A053 903 MARYLAND UNIV COLLEGE PARK COMPUTER SCIENCE CENTER F/G 12/1
CONNECTIVITY IN LATTICES AND MOSAICS.(U)
FEB 78 N AHUJA AFOSR-77-3271
UNCLASSIFIED TR-637 AFOSR-TR-78-0776 NL

1 OF 1
AD A
053903

END
DATE
FILMED
6-78
DDC


1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5
5·0
NATIONAL BUREAU OF STANDARDS

(2)

COMPUTER SCIENCE TECHNICAL REPORT SERIES

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release LAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

(2)

TR-637
AFOSR-77-3271

February 1978

CONNECTIVITY IN LATTICES AND MOSAICS

Narendra Ahuja
University of Maryland
Computer Science Center
College Park, MD 20742

ABSTRACT

This paper is concerned with connectivity in regular lattices and random mosaics, and makes two major contributions. First, it presents a solution to the hitherto unsolved problems of predicting the expected numbers of connected components in square and hexagonal lattices, using a one dimensional growth approach that can also be used for some other lattices. Second, it investigates the relationship between connectivity in regular lattices and random mosaics. Experimental results are presented for both the cases.

The support of the U.S. Air Force of Scientific Research under Grant AFOSR-77-3271 is gratefully acknowledged, as is the help of June Slack in preparing this paper. The author also acknowledges helpful conversations with Prof. E. Slud.

1. Introduction

In an earlier paper [1] we proposed mosaic models for images and presented an outline of our intended approach towards developing well-defined procedures for the application of these models in image modelling. Central to our scheme is the knowledge of the various features of the patterns generated by the different models. Two examples of such features relate to the properties of the connected components formed by identically colored neighboring units (adjacent cells or bombed areas). These features appear to be important for the following reasons:

1) In analyzing an image we cannot isolate a single unit. Consequently any known geometrical properties of the units [7] provided by the various models may not be of much help. The smallest available homogeneous, isolated region is a connected component.

2) For the same fractional coverage of the pattern by the given colors, different models are likely to provide patches having different topological and geometrical properties, such as convexity, elongatedness, dispersedness, etc. These properties of the individual components seem to be crucial in texture perception and have been included in most of the definitions of texture [3].

3) The relative area of a component, or, alternatively, the density of components, is a measure of the texture busyness.

This paper considers connectivity properties only for cell structure models. The approach is divided into two major steps:

a) Obtain the desired properties for a regular lattice that describes a regular graph which is the dual (in an expected sense, if necessary, as we shall see later) of the cell structure tessellation under consideration.

b) Interpret the properties of the components in the pattern generated by the model in terms of those pertaining to their counterparts in a regular lattice.

Section 2 illustrates our approach to obtaining the various desired properties for a regular square lattice. In Section 3 we obtain corresponding results for a regular hexagonal lattice. Although our method is fairly general and can be used for other lattices too, we will consider only the square and hexagonal lattices, for reasons that will be made clear later.

Section 4 relates the results for regular lattices to the corresponding properties of the appropriate random mosaics. Section 5 provides experimental results on the observed and predicted expected numbers of components in square and hexagonal lattices, and the Poisson line and Occupancy mosaics. In Section 6 we present some general comments on our method and suggest some future projects.

2. The Square Lattice

In this section we will discuss the problem of obtaining various properties of the black connected components in a pattern formed by coloring each point of a square grid black with probability p, and white with probability q = 1-p.

2.1. Expected number of components (per m x m window)

This comparatively simple problem has interested theoretical physicists for many years, but has not yet been solved [5]. Roach [5] provides two estimates of lower and upper bounds on the number of components per m x m window. In one he considers certain features of the immediate neighborhood of a black point, which are not very common among the points belonging to a component, but which occur for at least one point per component. The probability that any arbitrary point has such a feature is readily computable. Roach estimates the number of components by estimating the expected number of points possessing such a feature. Since this, in general, is an overestimate, as remarked above, he successively refines the estimates by counting the exceptions that make the estimates erroneous.

In his second approach Roach considers the various configurations of black points that can give rise to a single component. The probability that a given point belongs to a given configuration is easy to compute. By considering the different shapes that a component of size n can have, the probability that a black point belongs to a component of size n can be computed. He estimates the number of components by computing the lower and upper bounds on the numbers of components of sizes 1 to 7. Domb et al. [2] improve Roach's bounds by considering the components of sizes 1 to 9.

The validity of the above estimates is ultimately limited by the increasingly intractable combinatorics for the large-size components. Since large components become more and more prevalent as p increases, the inaccuracies of the estimates become increasingly pronounced, as shown in Fig. 1 [5].

The above approaches to component analysis are likely to be relatively complex as compared to viewing the components as the result of a growth process in two dimensions, which in turn should be harder than if one could interpret the image as having been produced by one dimensional growth. The combinatorial complexity then gets divided and there is a better hope of either deriving the explicit form of dependence, or, at least, confining the intractability to fewer parameters which, if known, would provide the solution. These parameters can be empirically estimated.

We now describe a specific approach involving empirical estimation of a single parameter.

Consider a white M x N image. Let each point of the first row be colored black with probability p and white with probability q = 1-p. The row now consists of a sequence of runs of black and white points. Any of these runs can have length between 1 and N. Let E(r) denote the expected number of runs of length r in a row. We shall first compute E(r).

A run of length r will occur if there are r consecutive black points, with the first and the rth points having a white neighbor. (Note that if a row-end is included in the run then the latter restriction is clearly unnecessary). We consider two cases:

Case (i) $0 < r < N$

It is clear that a run of length r can occupy one of (N-r+1) positions in the row. In all but two of these positions, the run must be surrounded by two white points, while in the remaining two positions it need have a white neighbor only on the interior end. Since each point is black with probability p, and white with probabliity q, we have

$$E(r) = (N-r-1)\, p^r q^2 + 2\, p^r q$$

Case (ii) $r = N$

The entire row should be black. Clearly

$$E(N) = p^N$$

Thus, the expected number of runs of length r per row is

$$E(r) = \begin{cases} (N-r-1)\, p^r q^2 + 2\, p^r q & 0 < r < N \\ p^N & r = N \end{cases}$$

As a result of coloring the first row, therefore, there are an expected number of E(r) potential components originating in the first row. We now start coloring the following rows. When the i^{th} row is considered each of its E(r) runs may be of one of two types:

1) It may be isolated from all of the runs in the $(i-1)^{st}$ row, i.e. none of the points in the run may be a neighbor of a black point in the $(i-1)^{st}$ row. Such a run may be the origin of a new component. Alternatively,

2) It may touch k, where $k \geq 1$, runs in the $(i-1)^{st}$ row, which may belong to ℓ different hitherto isolated components, $1 \leq \ell \leq k$. Each such run thus merges ℓ currently distinct components into one.

Since each point is independently colored, the probability that a run of length r in the i^{th} row does not overlap with any run in the $(i-1)^{st}$ row is simply q^r. We will now compute the expected number of runs that a given run of length r overlaps.

Each of the overlapping runs can be uniquely labelled by its left end. The probability that a point is a left end is the same as the probability that it is black and its left neighbor is white, which is qp. Since any point can be a left end independently of all the other points, the expected number of left ends among r points is given by rqp.

In order to compute how many runs a run of length r overlaps, note first that the point directly below the left end of the run does not have to be a left end. Since its left neighbor has no neighbor among the points in the run, its color is immaterial. Thus, the expected number of runs that a given run of length r overlaps is $(r-1)qp + p$.

One or more of these overlapping runs may belong to the same component. Moreover, the runs belonging to the same component need not be adjacent. It is this characteristic of the growth that makes the influence of the next row on the status of connectedness depend upon the entire past (all the previous rows). However, this single, relatively simple characteristic of the pattern embodies the entire combinatorial complexity which manifests itself in many ways in other representations. Given this parameter, the process that determines the number of connected components becomes Markov of order 1.

Let T(r) be the expected number of currently distinct components reaching the $(i-1)^{st}$ row, whose runs are overlapped by a run of length r in the i^{th} row. Then the increment in the number of components as a result of the addition of a single row, Δ, is given by

$$\Delta = \sum_{r=1}^{N} \text{(Expected number of runs of Length r)[Pr (a run of length r is isolated) + Pr (a run of length r is not isolated) (1 - T(r))]}.$$

We have divided the runs into two categories - isolated and nonisolated - for the following reason. T(r) is an expected value and adds further to the variance of the final results. We would like to minimize this effect. Estimating the probability distribution of the number of components overlapping a run of length r, however, becomes very cumbersome. Since the probability of a run being isolated or not is easy to find, we confine the impact of T(r) only to the nonisolated runs, by estimating and using its value only for such runs. Thus,

$$\Delta = \sum_{r=1}^{N} E(r) \quad [q^r + (1-q^r) \ (1-T(r))]$$

If C_o denotes the expected number of components in the first row, then the expected number of components, C, in an M x N image can be expressed as

$$C = C_o + (M-1) \ \Delta$$

We can approximate C_o by the expected number of runs in a row. Then,

$$C_o = (N-1)\ qp + p$$

so that $$C = (N-1)\ qp + p + (M-1)\ \Delta$$

The above estimate of C_o, however, is an overestimate and should be refined, as we will see in Section 5.

2.2. Expected area of a component

When each point of the image is colored black with probability p, we have

Expected number of black points in the image = pMN.

We already know the expected number of components, C, in the image. Therefore

Expected area of a component $\equiv A = \frac{pMN}{C}$.

2.3. Expected perimeter of a component

We will define the perimeter of a component as the number of pairs of points a and b such that a is a black point belonging to the component, and b is a white neighbor of a [6]. The reason for using this definition will be made clear in Section 4.

We will first compute the total length of perimeter in the image.

Clearly,

$$\Pr\{\text{A black point has j white neighbors}\} = \binom{4}{j} (1-p)^j\ p^{4-j}$$

Then,

$P_1 \equiv$ Expected number of white neighbors of a black point

$\equiv$ Expected contribution to the total perimeter due to a single black point

$$\equiv \sum_{j=1}^{4} j \binom{4}{j} (1-p)^j\ p^{4-j} = 4(1-p)$$

The expected perimeter of a component, then, is given by

$$P = \frac{pMN}{C}\ P_1 = AP_1$$

2.4. Expected dispersedness of a component

We can use the ratio of the square of the perimeter P of a component to its area as a measure of its dispersedness [6]. We have obtained an expression for the expected perimeter of a component. Hence the expected dispersedness of a component is given by

$$\frac{P^2}{A} = \frac{A^2 P_1^2}{A} = AP_1^2$$

Experimental details and results using this estimate are provided in Section 5.

3. The Hexagonal Lattice

We would now like to obtain the properties of connected components in a hexagonal lattice. A hexagonal lattice can be viewed as a square lattice whose alternate rows of points have been displaced horizontally by half the interpoint distance (Fig. 2). The number of points per row of a given picture now may be different if the picture has an odd number of columns. A row, then, will have either of two lengths that differ by 1. This will change the run statistics of the rows. Since there are two kinds of rows, we can evaluate Δ for each kind, and evaluate their contributions to the expected total number of components in the picture according to the numbers of the two kinds of rows, which will again differ by 0 or 1, depending on whether the image has an even or odd number of rows. For a large image, however, these end effects may be ignored. We will give expressions which take into account these effects.

3.1. Expected number of components

Let us assume that the even-numbered rows have been shifted with respect to the odd-numbered rows. Let Δ_o and Δ_e denote the value of Δ for odd and even numbered rows, respectively.

For a picture with N columns the expected number of runs of length r in an odd numbered row is given by

$$E_o(r) = \begin{cases} (\lceil \frac{N}{2} \rceil - r - 1)\, p^r q^2 + 2p^r q & 0 < r < \lceil \frac{N}{2} \rceil \\ p^r & r = \lceil \frac{N}{2} \rceil \end{cases}$$

and the expected number of runs of length r in an even numbered row is given by

$$E_e(r) = \begin{cases} (\lfloor \frac{N}{2} \rfloor - r - 1) p^r q^2 + 2p^r q & 0 < r < \lfloor \frac{N}{2} \rfloor \\ p^r & r = \lfloor \frac{N}{2} \rfloor \end{cases}$$

Unlike the square lattice case the probability that a run in a hexagonal lattice row depends upon its location in the row, and the type of the row. A row-interior run of length r overlaps exactly r+1 points in the previous row. If the run includes one of the row ends it is connected to r points in the previous row. A completely black row is isolated if the previous row is white. Thus,

$I(r) \equiv$ Pr (a run of lenght r is isolated)

$=$ Pr (the run is an interior run) q^{r+1} +

Pr (the run includes exactly one row end) q^{r} +

Pr (the run consists of the entire row) $\cdot\, q^{\{\text{length of the previous row}\}}$

Now, for $0 < r <$ row length

$$\Pr(\text{a run is a row-interior run}) = \frac{\text{Number of interior positions of a run in a row}}{\text{Total number of possible positions of a run in the row}}$$

$$\Pr(\text{a run includes one row end}) = 1 - \Pr(\text{the run is a row-interior run})$$

If $I_o(r)$ and $I_e(r)$ denote the values of $I(r)$ for the odd and even numbered rows, respectively, it is clear that

$$I_o(r) = \begin{cases} \dfrac{\lceil \frac{N}{2} \rceil - r - 1}{\lceil \frac{N}{2} \rceil - r + 1} q^{r+1} + \dfrac{2}{\lceil \frac{N}{2} \rceil - r + 1} q^{r} & 0 < r < \lceil \frac{N}{2} \rceil \\ q^{\lfloor \frac{N}{2} \rfloor} & r = \lceil \frac{N}{2} \rceil \end{cases}$$

and

$$I_e(r) = \begin{cases} \dfrac{\lfloor \frac{N}{2} \rfloor - r - 1}{\lfloor \frac{N}{2} \rfloor - r + 1} q^{r+1} + \dfrac{2}{\lfloor \frac{N}{2} \rfloor - r + 1} q^{r} & 0 < r < \lfloor \frac{N}{2} \rfloor \\ q^{\lceil \frac{N}{2} \rceil} & r = \lfloor \frac{N}{2} \rfloor \end{cases}$$

The expected value of the increment in the number of components due to an odd numbered row is

$$\Delta_o = \sum_{r=1}^{\left\lceil \frac{N}{2} \right\rceil} E_o(r) \left\{ I_o(r) + (1-I_o(r)) \ (1-T(r)) \right\}$$

and

$$\Delta_e = \sum_{r=1}^{\left\lfloor \frac{N}{2} \right\rfloor} E_e(r) \left\{ I_e(r) + (1-I_e(r)) \ (1-T(r)) \right\}$$

The expected total number of components in the image with M rows is then

$$C = C_o + \left\lceil \frac{M-1}{2} \right\rceil \Delta_e + \left\lfloor \frac{M-1}{2} \right\rfloor \Delta_o$$

where $C_o \equiv$ expected number of components in the first row.

3.2. Expected area of a component

Expected number of black points in the image $\equiv B =$

$$p\left[\left\lceil \frac{M}{2} \right\rceil \left\lceil \frac{N}{2} \right\rceil + \left\lfloor \frac{M}{2} \right\rfloor \left\lfloor \frac{N}{2} \right\rfloor \right]$$

Therefore,

Expected area of a component $\equiv A = \frac{B}{C}$.

3.3. Expected perimeter of a component

Using the same notation as used for the square lattice in Section 2.3, we have

$$P_1 = \sum_{j=1}^{6} j \binom{6}{j} (1-p)^j \ p^{6-j} = 6(1-p)$$

The expected perimeter of a component is given by

$$P = AP_1$$

3.4. Expected dispersedness of a component

The expected dispersedness of a component is given by

$$\frac{P^2}{A} = AP_1^2$$

4. Regular Lattice Patterns and Cell Structure Mosaics

4.1. Regular tessellations

Our motivation for the connectedness analysis of regular lattice patterns is to deduce from it the corresponding results for certain cell structure mosaics.

Definition:

Let T be a tessellation of plane into cells. Let any two cells be called neighbors if they share an edge. We will call T a KV regular tessellation if each of its cells has exactly K neighbors and exactly V cells meeting at each of its vertices.

Let $F_{p,T}$ denote the probability distribution of the number of connected components obtained as a result of coloring each of the cells of T black with probability p, and white with probability 1-p.

Theorem:

If T_1 and T_2 are two KV regular tessellations, for any positive integers K and V, then for any $0 \leq p \leq 1$ we have

$$F_{p,T_1} \equiv F_{p,T_2}$$

Proof:

We will view a connected component as the result of a growth process. At any stage during the growth we label a component by (a) its size S, i.e. the number of cells in it, and (b) the number of unexamined (uncolored) neighbors of the component, denoted by U. Clearly, in the beginning when the component has only one cell, S = 1 and U = K.

At any step during the growth, the unexamined neighbors of a component are colored one by one until one is colored black. The size of the component then increases by 1 and the process repeats. If at any step all of the U unexamined neighbors are eventually colored white, the component cannot grow any further. Thus the size of the component, or the number of steps taken before the growth stops, depends on the lattice only through the probability distribution of the number of unexamined neighbors at different steps in the growth.

Now, by definition, a KV regular tessellation describes a completely regular graph [4]. The dual graph of a KV regular tessellation describes the connectedness among cells, and, by definition of complete regularity, is also completely regular. This means that the dual graphs of all KV regular tessellations are isomorphic to a single, representative completely regular graph. Clearly, any two such graphs would result in identical probability distributions of the number of unexamined neighbors at different steps. Therefore the growth processes will be identical in all KV regular tessellations. In particular the probability distribution of the size of a connected component in any two KV regular tessellations T_1 and T_2 will be identical, i.e.

$$F_{p,T} \equiv F_{p,T_2}$$

This completes the proof.

By a straightforward application of Euler's polyhedral formula it can be shown [4] that there exist only three repetitive planar graph patterns, or mosaics; and they can be formed

by using triangles, quadrangles, or hexagons as the primitive repeating pattern. The K and V values of the corresponding KV regular tessellations, and their types, are given in Table 1.

Table 1

K	V	Resulting Tessellation
3	6	Triangular
4	4	Quadrangular
6	3	Hexagonal

4.2. Random mosaics

We now consider the connectedness problem for cell structure random mosaic models. Some of these models have the property that in the corresponding tessellations, the expected values of K and V are the same for all the cells. For the Poisson line and the occupancy models, these values are the same as those for the quadrangular and hexagonal tessellations, respectively. One may expect, therefore, that the growth processes in the above tessellations and the corresponding mosaics would be similar in characteristics. However, without having a complete mathematical description of the growth processes, e.g. the probability distribution of the number of unexamined neighbors at various steps, we cannot prove exactly which of the characteristics will be identical for the two processes. However, we may expect that the equal number of positive and negative variations in the values of K and V for different cells in large Poisson line

or occupancy model tessellations may cancel out, and the expected area of a component in the mosaics may be the same as the expected area of a component in the patterns based upon the corresponding KV regular tessellations. We then have the following conjecture.

Conjecture:

Let T_1 be a KV regular tessellation. Let T_2 be a tessellation each of whose cells has an expected number K of neighbors, and an expected number V of cells meeting at each of its vertices. Let I_i denote the patterns obtained by coloring the cells of T_i black with probability p, and white with probability 1-p. Then the expected area of a black component in I_1 is the same as in I_2.

Experimental observations on the expected number of components in the square lattice, Poisson line, hexagonal lattice, and occupancy mosaics, are presented in Section 5.

In our current repertoire of cell structure models there are none known to have tessellations corresponding to the triangular tessellation. On the other hand, the Poisson line and occupancy models, as mentioned earlier, do correspond to the quadrangular and hexagonal tessellations. This is the reason why, in Section 3, we carried out the connected component analysis only for the latter two lattices. The triangular lattice analysis, if needed, can also be carried out along the same lines.

4.3. Estimation of patch properties

The analysis presented above predicts the expected number of cells in a connected component, or a patch, in the mosaic.

The expected values of the properties of cells in some tessellations are also known [7]. We provide below the expressions for the expected patch area and perimeter, in terms of the corresponding properties of the connected components as obtained above.

Expected patch area = (Expected size of a component) x
(Expected area of a cell)

The perimeter of a patch is formed from a sequence of cell edges. None, some or all of the edges of a black cell could belong to the perimeter of a patch. This requires that to compute perimeter we must consider all the edges of a cell that border a white neighbor. The definition of perimeter that we chose for lattice components in Section 2 extends meaningfully to patches in the above sense. We have

$$\text{Expected patch perimeter} = (\text{Expected perimeter of a component}) \times \frac{\text{Expected cell perimeter}}{\text{Expected number of sides of a cell}}$$

5. Experimental Results

In this section we provide the details of the experiments carried out on the regular lattices as well as on the random mosaics. The predictions and verifications were made only for the expected number of connected components in the image.

5.1. The regular square lattice

As pointed out in Section 2, we must estimate the parameter T(r) for the lattice in order to be able to predict the expected number of connected components in a given image. However, we should note that for a finite size lattice the number of connected components in the lattice would be influenced by the border. The expected number would be overestimated due to the presence of distinct components touching the border that could have merged had the picture extended still further. Any estimate of T(r) will, therefore, clearly, be a function of not only the length r of the run, but also of where the run occurs relative to the image border. To avoid this complexity we will restrict our analysis only to the interior of the lattice where the border effects have presumably died down (this assumption becomes less and less true as the probability p of a point being black approaches 1).

To estimate T(r) we observed its values for all runs that occured in a 100 x 100 window located at the center of a 500 x 500 image. Since the probability of a run being of length r decreases monotonically with r($0<p<1$), the observed number of runs of length r also decreases with r. We estimated T(r) only for those $r \leq R$ such that

a) The number of nonisolated runs of length greater than R is no more than 5% of the total number of nonisolated runs, and

b) T(r) is monotonic for $r \leq R$ (this is to reduce sensitivity to erratic fluctuations in the value of T(r) that might result from unavailability of a sufficient number of long runs in the chosen, finite window). For all the runs of length $r>R$, we observe a cumulative mean value of T(r), denoted by T(R+1).

The above observations were made on ten different pictures for $p = .1,\ldots,.9$, in steps of .1. For each value of p, the number of connected components in the window was also observed. Those pictures giving rise to extreme values of this number are discarded, and the remaining eight are used to estimate T(r)'s. The results are shown in Table 2.

The estimated values of T(r) were also used to predict the expected number of components in a 100 x 100 image. As noted in Section 2, the expression provided there for C_o is an overestimate of the true value of C_o, since it estimates the number of different components in a row by the number of runs in a row. Similarly, the number of isolated runs in the initial row will be an underestimate of the number of different components in the row. However, if the picture has a large number of rows, the influence of the errors in the two estimates on the number of connected components in the image will decrease. In terms of the notation of Section 2 we compute C_o as follows:

$$C_o = \frac{(N-1)qp + p + (M-1)\Delta}{M}$$

$$= \frac{(\text{Number of runs in the first row}) + (M-1)\Delta}{M}$$

Table 2

p	R	Values* of T(r)-1 for r ≤ R+1								
		1	2	3	4	5	6	7	8	9
.1	2	0.0	0.0	.0417						
.2	2	0.0	0.0	.0779						
.3	3	0.0	0.0	.1007	.2863					
.4	3	0.0	0.0	.1025	.4100					
.5	4	0.0	0.0	.1059	.2922	.6578				
.6	6	0.0	0.0	.0942	.2282	.3773	.5501	.8293		
.7	8	0.0	0.0	.0537	.1189	.1778	.2387	.2773	.3487	.4827
.8	6	0.0	0.0	.0131	.0384	.0458	.0667	.1002		
.9	2	0.0	0.0	.0095						

*Note that T(R+1) is the cumulative value of T(r) for all r>R.

The observed and the predicted expected numbers of components in a 100 x 100 window are shown in Table 3. We may note that the errors between the observed and the predicted values remain much smaller than the differences between the lower and the upper bounds shown in Fig. 1. However, the errors do increase with increasing p. This may be expected in view of our earlier remarks that given a fixed window, the number of components in it becomes more and more influenced by the border effects as the value of p approaches 1.

Table 3

p	Expected number of Components		Estimated* standard deviation of the observed values $\hat{\sigma}$	b-a	$\frac{\lvert b-a \rvert}{\hat{\sigma}}$
	Observed (a)	Predicted (b)			
.1	798.37	801	16.40	2.62	.16
.2	1215.25	1220	13.90	4.75	.35
.3	1291.25	1289	23.15	-2.25	.1
.4	1085.25	1074	11.96	-11.25	.97
.5	673.75	678	16.55	4.25	.26
.6	262.62	267	11.87	4.38	.37
.7	72.62	62	6.74	-10.62	1.59
.8	13.25	18	3.10	4.75	1.54
.9	1.88	4	.93	2.12	2.16

* Unbiased estimate almost the same as the maximum likelihood estimate for the number of data points n = 10.

5.2. The Poisson line mosaic

In Poisson line mosaics formed by digital straight lines there is a nonzero probability of a small cell being lost due to the finite thickness of the lines. In other words, what would be a cell with small interior area and closely spaced borders in a Poisson line mosaic in the Euclidean plane, may appear in the digital mosaic as overlapping and/or neighboring digital straight lines with no points belonging to the interior of the cell. Such cells, if not properly taken care of, would change the connectedness properties of the digital Poisson line mosaic.

One way to reduce this effect is to decrease the probability of a cell being very small. This can be achieved by using a Poisson process of low intensity, i.e. having fewer straight lines per unit area of picture. However, this will also reduce the number of cells per unit area, and thus, in order to have a sufficiently large number of cells in the mosaic for any meaningful conclusion about its expected connectedness properties, we will have to work with larger images.

At this stage we may point out that our aim is to investigate the conjectured similarity of the connectedness properties of the points in a square lattice and the cells in a Poisson line mosaic. If we decide to do it by comparing the expected number of connected components per m x m window in the interior of the image to reduce border effects, we may have to use an inconveniently large Poisson line mosaic. However, if the conjecture is true the border effects on the expected number of connected components in a Poisson line mosaic of size M x M, having a sufficiently large number of cells, say N, should be identical

to those in a $\sqrt{N}$ x $\sqrt{N}$ square lattice pattern. Thus we may only have to compare the observed number of connected components in an M x M mosaic with that of an $\sqrt{N}$ x $\sqrt{N}$ square lattice. This is the approach that we have taken.

We generated a Poisson line tessellation of size 1000 x 1000, using a Poisson process of an appropriate intensity so that the number of cells in the resulting image was 100. This was done to make the comparison with a square lattice easier. However, in spite of the small number of cells, some of them were small enough, as a result of large cell area variance, to be lost on account of the finite thickness of the digital lines. In order to avoid the use of still larger size mosaics we chose to work with the Euclidean plane version of the same mosaic. The cells were colored by hand and the resulting number of components counted on three independent sets of patterns, obtained from the same tessellation. For the corresponding square lattice connectivity results, we observed the number of connected components on a set of ten 10 x 10 images. Both the mosaic as well as the square lattice had 100 units (cells and points), so that the results for the two cases did not need to be normalized for scale.

The average observed number of connected components on the 10 images should provide an estimate of the expected number of connected components in a 10 x 10 square lattice. If the conjecture is true this should also be the estimate of the expected number of connected components in the mosaic.

To test the validity of the conjecture, we must investigate whether the average observed number of components in the mosaic is significantly different from that for the corresponding square lattice case. If this is so, then the hypothesis is false. If not, then the data do not falsify the conjecture. In view of the lack of any knowledge about the probability distribution of the number of connected components in the mosaic, and the limited amount of data, we have only computed the following measure of the deviation of the average observed number of components in the mosaic from the value predicted by the conjecture.

Let A ≡ Number of connected components in a 10 x 10 square lattice.

B ≡ Number of connected components in a mosaic consisting of 100 cells.

$\bar{X}$ ≡ Average value of X, for any X.

σ_X ≡ Variance of X.

s_X ≡ Sample variance of X.

Then, $z_B = \frac{|\bar{B}-\bar{A}|}{\sigma_B}$ is a normalized measure of the deviation of the expected number of connected components in the mosaic from that in the square lattice. z_B should have a small value if the conjecture is true.

Since we do not know σ_B, we will estimate it from the observed sample variance, s_B. The formula $\hat{\sigma}_B = \sqrt{\frac{s_B}{n-1}}$ provides an unbiased estimate of σ_B based upon n observations of B. Table 4 lists the observed values and the computed measures of the deviations. To be more conservative we could use the maximum likelihood estimate of σ_B, which would increase the deviations by a factor of $\sqrt{\frac{n}{n-1}}$. This evaluates to 1.25 for n = 3.

Table 4

p	$\bar{A}$ Computed from ten images	Observed values of B B_1	B_2	B_3	$\bar{B}$	$\hat{\sigma}_B$ *	$\bar{B}-\bar{A}$	$z_B = \frac{\lvert\bar{B}-\bar{A}\rvert}{\hat{\sigma}}$
.1	8.00	8	10	7	8.33	1.55	.33	.21
.2	14.20	12	20	12	14.66	4.60	.46	.10
.3	14.50	15	12	14	13.66	1.55	-.84	.55
.4	12.10	12	5	13	10.00	4.36	-2.10	.48
.5	8.90	9	11	7	9.00	2.00	.10	.05
.6	6.20	3	5	6	4.66	1.55	-1.54	1.00
.7	3.10	3	2	2	2.33	.58	-.87	1.50
.8	1.60	2	1	4	2.33	.58	.73	1.26
.9	1.00	1	1	2	1.33	.58	.33	.58

* Unbiased estimate

However, for both the estimates of σ_B the normalized deviations remain small, and thus, the data do not disprove the conjecture.

5.3 The regular hexagonal lattice

As in the square lattice case we require an estimate of T(r) in order to predict the expected number of components in a given pattern. It is clear from the topology of the lattice (Fig. 2) that we need an image with 500 rows and 1000 columns in order to work with the same amount of data as in a 500x500 square lattice. The values of T(r) were observed for a 100 row by 200 column window at the center of the 500x1000 image. For the reasons explained in Section 5.1, the values of T(r) were obtained only for $r \leq R$. For all $r > R$, a cumulative mean value of T(r) was used, denoted by T(R+1).

Due to high computational costs the estimates were obtained from only seven different images, for each of the values of p =.1,...,.9, in steps of .1. The number of connected components in the window was also obtained in each case, and the data from those pictures having the extreme numbers of components was not used in estimating T(r)'s. Table 5 shows the results. In accordance with the analysis of Section 3, these values of T(r) were then used to predict the expected number of components in a 100x200 image. The predicted and the observed values of the expected number of components, and their normalized differences, are shown in Table 6.

Table 5

p	R	Values* of $T(r)_{-1}$ for $r \leq R+1$						
		1	2	3	4	5	6	7
.1	2	0.0	.0291	.1333				
.2	2	0.0	.0765	.1710				
.3	3	0.0	.0828	.1909	.3695			
.4	4	0.0	.1005	.2196	.3663	.7073		
.5	5	0.0	.0868	.1949	.3448	.5028	.7172	
.6	6	0.0	.0451	.1000	.1467	.1995	.2083	.3067
.7	4	0.0	.0090	.0323	.0479	.0926		
.8	2	0.0	.0023	.0116				
.9	1	0.0	.0011					

*Note that T(R+1) is the cumulative value of T(r) for all r > R.

Table 6

p	Expected number of Components		Estimated* standard deviation of the observed values $\hat{\sigma}$	a-b	$\frac{\lvert b-a \rvert}{\hat{\sigma}}$
	Observed (a)	Predicted (b)			
.1	731	722	4.60	9.0	1.95
.2	969.2	967	7.41	2.2	.30
.3	869.0	865	27.58	4.0	.15
.4	559.2	538	20.84	21.2	1.02
.5	184.8	184	8.80	0.8	.09
.6	40.4	52	4.84	-11.6	2.40
.7	8.6	8	2.06	0.6	.29
.8	2.0	5	.63	-3.0	4.74
.9	1.0	1	0	0.0	0.0

*Unbiased estimate. Almost the same as the maximum likelihood estimate for the number of data points n = 10.

5.4 The occupancy mosaic

We generated three occupancy mosaics having 100 cells each, by using 100 nuclei for each tessellation. On account of the obvious differences in the nature of the underlying geometrical processes, the thickness of the digital lines in a digital occupancy mosaic is less likely to cause loss of cells than it was in the case of the Poisson line mosaic. The smaller cell area variance allows a larger number of cells in an image of given size. For this reason, an image size of 218x218 was found sufficient to accommodate 100 cells. The component counting was done automatically on the three mosaics obtained from three different tessellations. (Due to the large-size Euclidean mosaics in the Poisson line case, and the involved manual work, we used the same tessellation to obtain three different mosaics in Section 5.2.) The results are shown in Table 7. The deviations, in general, remain small.

5.5 General remarks

Before closing this section we would like to present some analytical comments on the results of the experiments. First, it would be desirable to obtain the estimates of $T(r)$ from larger images and more observations than we considered. Although the accuracy of the estimates will always suffer from lack of data due to the limiting behavior of the patterns for $p = 1$, by using increasingly larger images and numbers of observations, we could improve

Table 7

p	$\bar{A}$ Computed from ten images	Observed values of B B_1	B_2	B_3	$\bar{B}$	$\hat{\sigma}_B$ *	$\bar{A}-\bar{B}$	$z_B = \frac{\lvert\bar{A}-\bar{B}\rvert}{\hat{\sigma}_B}$
.1	8.60	8	6	7	7.0	.82	1.60	1.96
.2	11.80	12	11	9	10.67	1.25	1.13	.91
.3	11.70	7	13	8	9.33	2.62	2.37	.90
.4	10.20	9	10	11	10.0	.82	.20	.24
.5	7.20	6	3	6	5.0	1.41	2.20	1.56
.6	4.10	5	2	4	3.67	1.25	.43	.35
.7	2.70	1	1	5	2.33	1.89	.37	.19
.8	2.0	2	2	1	1.67	.47	.33	.71
.9	2.0	1	1	1	1.0	0.0	-1.0	--

*Unbiased estimate.

the estimates for increasing values of p. This is true because we are estimating only the values of T(r), and not the nature of its dependence on p. The size of the data in our experiments was restricted by the associated computational costs. Because of the greater connectivity of the hexagonal lattice (occupancy mosaic) as compared to the square lattice (Poisson line mosaic) the above remarks have more serious implications for the former. For the same amount of insensitivity of an interior window to the border as in the square lattice case, we need a larger image. In our observations the values of T(r) exhibited great variability for large values of p, and this was greater for the hexagonal lattice than for the square lattice. This may explain why the corresponding predictions are poorer.

We have tested our conjecture assuming the border effects to be the same in both the regular lattices and the corresponding random mosaics. It may be reasonable to expect that the number of border points (cells) determine the magnitude of border influence, for a given size lattice (mosaic). In the mosaics used in our experiments the observed numbers of border cells were 41 for the Poisson line mosaic, and 29, 30 and 33 for the occupancy mosaics. (The corresponding regular lattices both have 36 border points.) It is difficult to judge the significance of these deviations and their effect on the number of connected components. This further indicates the need for

large amounts of data in experiments of the kind we have been concerned with here.

6. Discussion and Related Projects

We have presented a one dimensional growth approach to the analysis of lattice patterns. Although the method does not yield closed form expressions for the properties of the connected components, it uses empirical data that can be obtained and tabulated, once and for all, for a given type of lattice. If desired the data may be represented by a set of best-fitting polynomials in p and r, and these polynomials may be used to replace T(r) in the formulas to yield closed form expressions. The resulting predictors do not then need any explicit specification of external, empirically observed data. The generality of our approach stems from the fact that the lattices of interest can often be laid out in the plane as a stack of rows of points with specified connectivity rules among the points, which, in general, may be different for different rows. The connected component properties can then be derived by the row incremental process as illustrated for the square and hexagonal lattices.

As we have seen, the errors in the predicted expected number of components are larger for larger p's. However, we may note that cases involving large p's may not be of great interest, since whenever the region of interest in a cell structure model occupies more than half the image area, we can interchange the roles of objects and background, without affecting the validity of the model. Hence we may need the results obtained only for $p \leq .5$, which are quite accurate.

The expected number of connected components in a given image starts decreasing with increasing probability, after reaching a maximum. Intuitively this is so because the distinct

components start coalescing. It may be of interest to determine if there exists a critical probability value p_c beyond which the entire set of black points begins to collapse rapidly into a single giant component. This would also be the probability at which the border effects start to become dominant.

We have measured the values of T(r) and verified the predicted expected number of components for the given picture and window sizes. For reasons mentioned in Section 5.5, the results will, in general, be biased due to border effects. Unless we estimate the functional form of the dependence of T(r) on p and on the picture size, this bias will always exist, although it could be made smaller by increasing the picture size. Since we do not know how the bias varies with picture size, we cannot specify how to select the picture size so as to reduce it below a certain predetermined value. Our choice of the picture and window sizes has been ad hoc, and although for low values of p the chosen sizes appear to be sufficient, it is not clear for what maximum value of p the results obtained are within reasonable tolerance and characterize only the lattices involved.

In view of the fact that there can exist at most three kinds of tessellations, as discussed in Section 4.1, the connectedness analysis of the triangular, square and hexagonal lattices, along with the conjecture of Section 4.2, should be sufficient to predict the connectivity of a cell structure mosaic based upon any conceivable model in terms of the properties of the cells generated by that model. The amount of variation in the connectivity properties would partly be

due to the variances of the cell properties. For example, we may expect the patch properties in the occupancy mosaics to possess smaller variances than in the Poisson line mosaics, because of the larger variances of the cell properties for the latter [7].

The problems that we have had with small cells, while proving the conjecture of Section 4 for the digital Poisson line mosaics, do not have any serious implications in terms of the applicability of the Poisson line model, since in using the model for a given image, we are going to be interested in the patches and not in the cells, which in any case have no independent existence. The intensity of the underlying Poisson process should be immaterial as long as the patches remain large enough so as not to be significantly distorted in shape due to the presence of an excessive amount of border.

For both the lattices and the mosaics, we have considered patterns of only two colors, black and white. The analysis, however, extends to patterns with any number of colors, since the pattern formed by each of the colors on the combined background of the rest is similar to the black-white case. In fact, with a larger number of colors the domainance of a single color may be less, and the corresponding black-white patterns may all correspond to low values of p. Consequently, there may not be any need to interchange the roles of any two colors in order to avoid the use of the less accurate connectedness results for large

values of p.

The approach described in this paper could also be used for the connectedness analysis of a class of bombing models. We will investigate this problem in a later paper.

REFERENCES

1. Ahuja, N., "Mosaic models for image analysis and synthesis", University of Maryland Computer Science T.R. 607, Nov. 1977.

2. Domb, E., and Sykes, M. F., "Cluster size in random mixtures and percolation processes", Physical Review, 122, 1961, p.77.

3. Lipkin, B. S. and Rosenfeld, A., eds. Picture Processing and Psychopictorics, Academic Press, New York, 1970.

4. Ore, Oystein, Graphs and their Uses, Random House, New York, 1963.

5. Roach, S., The Theory of Random Clumping, Methuen and Co. Ltd., London, 1968.

6. Rosenfeld, A. and Kak, A. C., Digital Picture Processing, Academic Press, New York, 1976.

7. Schachter, B. and Ahuja, N., "A survey of random pattern generation processes", University of Maryland Computer Science T.R. 549, July 1977.

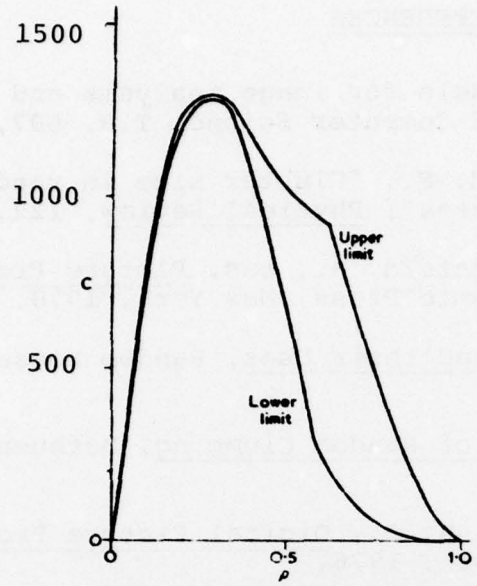


Fig. 1. Rigorous limits for the number of connected components on a square lattice of 1000 points [5].

Fig.2. A hexagonal lattice.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-78-0776
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)*: CONNECTIVITY IN LATTICES AND MOSAICS

5. TYPE OF REPORT & PERIOD COVERED: Interim rept.,

6. PERFORMING ORG. REPORT NUMBER: TR-637

7. AUTHOR(s): Narendra Ahuja

8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3271

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of Maryland
Computer Science Center
College Park, MD 20742

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F 2304 A2

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Office of Scientific Research/NM
Bolling AFB, DC 20332

12. REPORT DATE: Feb 78

13. NUMBER OF PAGES: 42 Pp.

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*

15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

image models
random geometry
connectivity
mosaics

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This paper is concerned with connectivity in regular lattices and random mosaics, and makes two major contributions. First, it presents a solution to the hitherto unsolved problems of predicting the expected numbers of connected components in square and hexagonal lattices, using a one dimensional growth approach that can also be used for some other lattices. Second, it investigates the relationship between connectivity in regular lattices and random mosaics. Experimental results are presented for both the cases.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

403 418